International Journal of Scholarly Research (IJSR)
Vol-1, Issue-2, 2017
http://www.scholars.org.in

"Review on Health Care Database Mining in Outsourced Database"

Monali Gainkar [1], Prof. Sonali Bodkhe [2]

[1] Departement of computer science and engg,
G.H. Raisoni Academy of engg. & technology, Nagpur
monalimgainkar@gmail.com
[2] Departement of computer science and engg,
G.H. Raisoni Academy of engg. & technology, Nagpur
sonali.bodkhe@raisoni.net

Abstract: Now A days there has been a outstanding growth in health data collection since the development of Electronic Medical Record (EMR) systems. Such collected data is further shared with others and analysed for Contrasting prospect. In contempt of much prosperity, data gathering and sharing have become a immeasurable concern as it menace individual privacy. In this paper, our goal is to propose a sheltered and secretive data management structure that addresses both the sheltered and secretive issues in the management or organization of medical data in outsourced databases. The proposed framework will assure the security of data by using semantically secure encryption schemes to keep data encrypted in outsourced databases. The framework also provides a differentially-private query or uncertainty interface that can support a number of SQL queries and complicated data mining responsibilities. For Private or without publically access of outsource data we are proposing a rule or protocol which is based on two secure multi-party Algorithm One for computing the union (or intersection) of private subsets that each of the interacting and collaborating players hold. Another is a protocol that tests the inclusion of an element held by one player in a subset held by another. So that all the purpose is to make a secure and private management system for medical data or record storage and accesses.

Keyword: Outsourced database, data mining, multi-party algorithm, data encryption.

1. INTRODUCTION

Healthcare industries store a massive amount of sensitive personal data, such as patient names, dates of birth, and personal medical records. Since healthcare data doubling every year, organizations need to invest in both hardware and software to store and manage large amount of data. Database outsourcing has gained importance in the past few years due to the emergence of the cloud computing. In Database-as-a-Service (DaaS), which is a sort of cloud computing services, the database proprietor outsources both databases and querying services to a cloud server and clients issue queries over the database to the cloud server. In this context, privacy is a most important test and it is necessary to satisfy main privacy requirements of database owners and clients. In the budding cloud computing archetype, data owners become progressively more aggravated to outsource their complex data management systems from confined sites to the commercial public cloud for great elasticity and financial savings. For the contemplation of users' privacy, susceptible data have to be encrypted before outsourcing, which makes valuable data utilization a very tough task. In this domain, cloud computing is an effective solution for healthcare companies to handle huge amounts of medical records. However, healthcare organizations face two technical challenges. First, data outsourcing exposes sensitive healthcare data to un-trusted cloud service providers. Unauthorized access to sensitive medical records can have a significant negative impact on healthcare services. To ensure the confidentiality of the medical data stored on the cloud, we should depend on semantically-secure encryption schemes. Using semantically-secure encryption schemes, it must be infeasible for a computationally-bounded adversary to derive significant information about a message when given only the cipher-text and the corresponding public key. In this regard, the challenge is how to ensure data confidentiality while allowing query execution over encrypted data.

Second, driven by mutual benefits and regulations, there is a demand for healthcare organizations to share patient data with various parties for research purposes. Healthcare organization may allow data analysts (e.g., researchers) to execute aggregate queries and perform some data analysis tasks (e.g., classification analysis) on the database. In this regard, the challenge is how to support aggregate queries or complex data mining tasks on encrypted data while preventing inference attacks.

There have been a lot of research proposals that separately address these two challenges. Most of the previous proposals on secure outsourced databases suggest encrypting the data before moving it to the cloud. While encryption can provide data confidentiality, it is of little use in deterring

inference attacks. Similarly, there is an extensive literature on private data analysis. However, all these proposals require access to unencrypted data to generate privacy-preserving answers and therefore do not satisfy the data confidentiality requirement. This reality demands a new privacy-enhancing technology that can simultaneously provide data confidentiality against an un-trusted database server, and prevent inference attacks from data analysts. We are proposing a general framework for secure and private data management in order to support effective data mining. The contributions of the our approach are summarized as follows:
Based on real-life healthcare scenarios, we will first identify a new problem of secure and private data management of outsourced databases for data mining purposes

- We adopt a new privacy-enhancing protocol that can provide data confidentiality against an un-trusted cloud server by using semantically-secure encryption schemes. We then extend the protocol to support aggregate queries or complex data mining tasks on encrypted data while preventing inference attacks.

- Taking decision tree learning as an example, we will show that it is possible to compute a classifier on the encrypted data. The computed classifier provides differential privacy guarantee to prevent an inference attack.

2. RELATED WORK

Previous work in privacy preserving data mining has considered two related settings. One, in which the data owner and the data miner are two different entities, and another, in which the data is distributed among several parties who aim to jointly perform data mining on the unified corpus of data that they hold.

F. McSherry [1] has developed Data records that are protected from the data miner. Hence, the data owner aims at anonym-zing the data prior to its release. He describe main approach in this context is to apply data perturbation. His developed idea is that the perturbed data can be used to infer general trends in the data, without enlightening original or secrete record information or data.

Lindell and Pinkas [2] has the goal is to perform data mining while protecting the data records of each of the data owners from the other data owners. This is a problem of secure multi-party computation. The usual approach here is cryptographic rather than probabilistic. He showed how to securely build an ID3 decision tree when the training set is distributed horizontally.

S. Barouti, D. Alhadidi and M. Debbabi [3] has faces the primary challenges such as privacy and it is most necessary to complete or satisfy main privacy of database owners and clients. He presents protocols for executing keyword search and combined SQL queries that maintain the confidentiality of both the client and the database owner. Client confidentiality is conserved such that the database owner and the cloud server cannot assume the constants contained in the query predicates. Database owner confidentiality is preserved such that the client cannot acquire any additional information or data accept the query result. The primitives that are utilized by them in designing these protocols include symmetric secret information retrieval and private integer assessment. They experimentally estimate the concert of the proposed protocols and report on the experimental results.

N. Cao, Z. Yang, C. Wang, K. Ren and W. Lou [4] for the first time ever, define and solve the dilemma of privacy-preserving query over encrypted graph-structured data in cloud computing (PPGQ), and establish a set of authoritarian privacy necessities for such a secure cloud data deployment system to become a reality. There work utilizes the standard of "filtering-and-authentication". They pre-established a feature-based index to provide feature-related information about each encrypted data graph, and then choose the well-organized interior product as the pruning tool to carry out the filtering route. They suggest a secure inner product working out technique, and then recover it to accomplish various privacy requirements under the known-background threat model.

F. Chen and A. X. Liu [5] has reflect on a two-tiered sensor network structural design in which cargo space nodes gather data from in close proximity sensors and answer queries from the descend of the network. The cargo space nodes provide as an intermediary tier between the sensors and the sink for storing data and dispensation queries. Storage nodes transport three main profits to sensor networks. First, sensors accumulate power by transfer all composed data to their neighbouring storage node as an alternative of sending them to the sink all the way through long routes. Second, sensors can be memory restricted because data are mainly stored on storage nodes. Third, query processing becomes more well-organized because the sink only converse with storage nodes for queries.

M. A. ALzain and E. Pardede [6] has suggest the design of a new reproduction suitable for NetDB2 architecture, known as NetDB2 Multi-Shares (NetDB2-MS). It is based on multi-service contributor and a undisclosed sharing algorithm instead of encryption, which is used by the previous NetDB2 service. They have done estimation through simulation and show a

considerable enhancement in performance for data storage and retrieval for various query types.

R. Mishra, D. P. Mishra [7] principally highlights some foremost security issues obtainable in current cloud computing surroundings. The primary issue that has to be dealt with when talking about data security in a cloud is protection of the data. There thought is to built a privacy preserving storehouse where data sharing services can modernize and organize the access and limit the usage of their shared data, as a substitute of put forward data to central establishment, and, hence, the repository will encourage data sharing and privacy of data. They aims at concurrently achieving data secrecy while still keeping the balancing relations intact in the cloud. His proposed system facilitates the data owner to assign most of computation intensive tasks to cloud servers without reveal data stuffing or user access right information.

3. PROPOSED METHOD

From the above effective and most precise discursion on previous work done by a great people about outsourcing the data in encrypted form in cloud for some most important benefits and studying and grabbing idea related to database query processing help to build our system.

Our Aim is to propose a sheltered and confidential data management structure that addresses. The proposed framework makes certain the security of data by using semantically-secure encryption system to keep data encrypted in outsourced databases. In this scenario the data in the form of datasets are stored in database act as a container and datasets are health care data which is very sensitive data that the data owner or organization of data doesn't want to reveal to next un-authorized party while outsourcing data in the cloud.

The framework also provides a differentially-private query interface that can maintain a number of SQL queries and difficult data mining tasks. In this scenario client can query to the server and apart from all this server can copied only the information or sensitive data to client only as per query held by him not more than that information is provided by the server. Because the data is very sensitive and this data kept in protection of the server side.

The Framework will address both the security and privacy issues in the management of medical data in outsourced database. In the face of many benefits, data collected works and allotment have become a big concern as it intimidate individual privacy. Our Idea is to propose a secure and private data management framework that addresses both the security and privacy issues in the management of medical data in outsourced database. The proposed framework ensures the security of data by using semantically-secure encryption schemes to keep data encrypted in outsourced databases. The framework also provides a differentially-private query interface that can support a number of SQL queries and complex data mining tasks.

For Private access of Outsource data we are proposing a protocol which is based on two secure multi-party Algorithm :

- One for computing the union (or intersection) of private subsets that each of the interacting players hold:

  In this we have a dataset of health care and we perform classification our datasets by making union of datasets or interaction of data sets and as a result of it the dataset we get is a common datasets and un-common datasets which is hold by a party or user which gone access this database.

- Another is a protocol that tests the inclusion of an element held by one player in a subset held by another:

  And after first step we are applying inclusion algorithm that can make a between the datasets held by one party with other one.

4. CONCLUSIONS

Our main goal is to make a system which provides a security and privacy preserving task in outsource database. An making query to database server and only that much information is provided to user. The proposed framework will ensures the security of data by using semantically-secure encryption schemes to keep data encrypted in outsourced databases of the health care database. This is more sensitive data for organization. The framework will also provide a differentially-private query interface that can support a number of SQL queries and complex data mining tasks. Our Multiparty Protocol will also guarantee the private access of outsourced data.

5. REFERENCES

[1] F. McSherry, "Privacy integrated queries," in Proceedings of the 35th ACM International Conference on Management of Data (SIGMOD), 2009.

[2] Y. Lindell and B. Pinkas, "Privacy preserving data mining," Journal of Cryptology, vol. 15, no. 3, pp. 177–206, 2002

[3] S. Barouti, D. Alhadidi, and M. Debbabi, "Symmetricallyprivate database search in cloud computing," in Cloud Computing Technology and Science (CloudCom), International Conference on, vol. 1. IEEE, 2013, pp. 671–678.

[4] N. Cao, Z. Yang, C. Wang, K. Ren and W. Lou, "Privacy-Preserving Query over Encrypted Graph-Structured Data in Cloud Computing," 2011 31st International Conference on Distributed Computing Systems, Minneapolis, MN, 2011, pp. 393-402.

[5] F. Chen and A. X. Liu, "SafeQ: Secure and Efficient Query Processing in Sensor Networks," 2010 Proceedings IEEE INFOCOM, San Diego, CA, 2010, pp. 1-9.

[6] M. A. ALzain and E. Pardede, "Using Multi Shares for Ensuring Privacy in Database-as-a-Service," 2011 44th Hawaii International Conference on System Sciences, Kauai, HI, 2011, pp. 1-9.

[7] R. Mishra, D. P. Mishra, A. Tripathy and S. K. Dash, "A privacy preserving repository for securing data across the cloud," 2011 3rd International Conference on Electronics Computer Technology, Kanyakumari, 2011, pp. 6-10.

[8] B. C. M. Fung, K. Wang, R. Chen, and P. S. Yu, "Privacypreserving data publishing: A survey of recent developments," ACM Computing Surveys, vol. 42, no. 4, pp. 1–53, June 2014.

[9] C. Dwork, "Differential privacy," in Proceedings of the International conference on Automata, Languages and Programming (ICALP), 2012.

[10] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in Proceedings of the 3rd conference on Theory of Cryptography (TCC), 2008.

[11] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed. Morgan Kaufmann Publishers Inc., 2011.

[12] M. D. Berg, O. Cheong, M. V. Kreveld, and M. Overmars, Computational Geometry: Algorithms and Applications, 3rd ed. Springer-Verlag TELOS, 2008.

[13] A. Frank and A. Asuncion, "UCI machine learning repository," 2010. [Online]. Available: http://archive.ics.uci. edu/ml

[14] Z. Zhu and W. Du, "Understanding privacy risk of publishing decision trees," in Proceedings of the 24th Annual IFIP WG

[15] Working Conference on Data and Applications Security and Privacy (DBSec), 2010.

[16] G. Jagannathan, K. Pillaipakkamnatt, and R. N. Wright, "A practical differentially private random decision tree classifier," Trans. Data Privacy, vol. 5, no. 1, pp. 273–295, Apr. 2012.